IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

N Vu
April 2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

Summary of Methodologies

- Data collected using SpaceX API, web scraping
- Data Wrangling
- Exploratory Data Analysis with SQL
- Exploratory Data Analysis with Data Visualization
- Interactive Visual Analytics with Folium
- Machine Learning Prediction

Introduction

- This capstone project helps determine “How much does a launch cost SoaceX?”. To keep costs down, SpaceX reuses its first stage, so finding the rate of successful launches would answer that.
- Various machine learning methods are used throughout this project to shed light on Falcon 9 launches and landings

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - gathered from SpaceX API and scrapped from Wikipedia
- Perform data wrangling
  - to clean the data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - to see which machine learning method works best

Data Collection

Data sets were collected by using

- SpaceX API (https://api.spacexdata.com/v4/rockets/)
  - Data provided is for various rockets owned by SpaceX that was then filtered out to only show Falcon 9 launches
- Missing values were replaced by mean values

Data Collection – SpaceX API

- With a focus on the Falcon 9 rocket, all launch records of that rocket were filtered out and used for this demonstration

- GitHub link: https://github.com/MelinoesGhost/Applied-Data-Science-Capstone-Falcon-X/blob/main/1.1%20jupyter-labs-spacex-data-collection-api.ipynb

Request data from SpaceX API (rocket launch data)

Decode response using .json() and convert to a dataframe using .json_normalize()

Request information about the launches from SpaceX API using custom functions

Create dictionary from the data

Create dataframe from the dictionary

Filter dataframe to contain only Falcon 9 launches

Replace missing values of Payload Mass with calculated .mean()

Export data to csv file

Data Collection - Scraping

- GitHub link: https://github.com/MelinoesGhost/Applied-Data-Science-Capstone-Falcon-X/blob/main/1.2%20jupyter-labs-webscraping.ipynb

Request data Falcon 9 launch data from Wikipedia

Create a Beautiful Soup object from HTML response

Extract column names from HTML table header

Collect data from parsing HTML tables

Create a dictionary from the data

Create dataframe from the dictionary

Export data to CSV file

Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts
- GitHub link: https://github.com/MelinoesGhost/Applied-Data-Science-Capstone-Falcon-X/blob/main/1.3%20labs-jupyter-spacex-Data%20wrangling.ipynb

Perform EDA and determine data labels

Calculate number of launches on each site

Calculate number of and occurrence of each orbit

Calculate number of and occurrence of mission outcome of the orbits

Create a landing outcome label from Outcome column

EDA with Data Visualization

Charts created

- Flight Number vs Launch Site (scatter point chart)
- Payload vs Launch Site (scatter point chart)
- Success rate of each orbit type (bar chart)
- Flight Number vs Orbit type (scatter point chart)
- Payload and Orbit type (scatter point chart)

- GitHub link: https://github.com/MelinoesGhost/Applied-Data-Science-Capstone-Falcon-X/blob/main/2.2%20jupyter-labs-eda-dataviz.ipynb.jupyterlite%20(2).ipynb

EDA with SQL

SQL queries performed:

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was achieved
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass. Use a sub-query
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.
- GitHub link: https://github.com/MelinoesGhost/Applied-Data-Science-Capstone-Falcon-X/blob/main/2.1%20jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Colored Markers of Launch Outcomes

- Added colored markers of successful (green) and unsuccessful (red) launches at each launch site to show which launch sites have high success rates

Distances Between a Launch Site to Proximities

- Added colored lines to show distance between launch site CCAFS SLC 40 and its proximity to the nearest coastline, railway, highway, and city

GitHub link:
https://github.com/MelinoesGhost/Applied-Data-Science-Capstone-Falcon-X/blob/main/3.1%20lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

Dropdown List with Launch Sites

- Allow users to select all launch sites or a certain launch site

Slider of Payload Mass Range

- Allow users to select payload mass range

Pie Chart Showing Successful Launches

- Allow users to see successful and unsuccessful launches as a percent of the total

Scatter Chart Showing Payload Mass vs. Success Rate by Booster Version

- Allow users to see the correlation between Payload and Launch Success

GitHub link:
https://github.com/MelinoesGhost/Applied-Data-Science-Capstone-Falcon-X/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- GitHub link: https://github.com/MelinoesGhost/Applied-Data-Science-Capstone-Falcon-X/blob/main/4.1%20SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Create NumPy array from the Class column

Standardize the data with StandardScaler; fit and transform the data

Split the data using train_test_split

Create a GridSearchCV object with cv=10 for parameter optimization

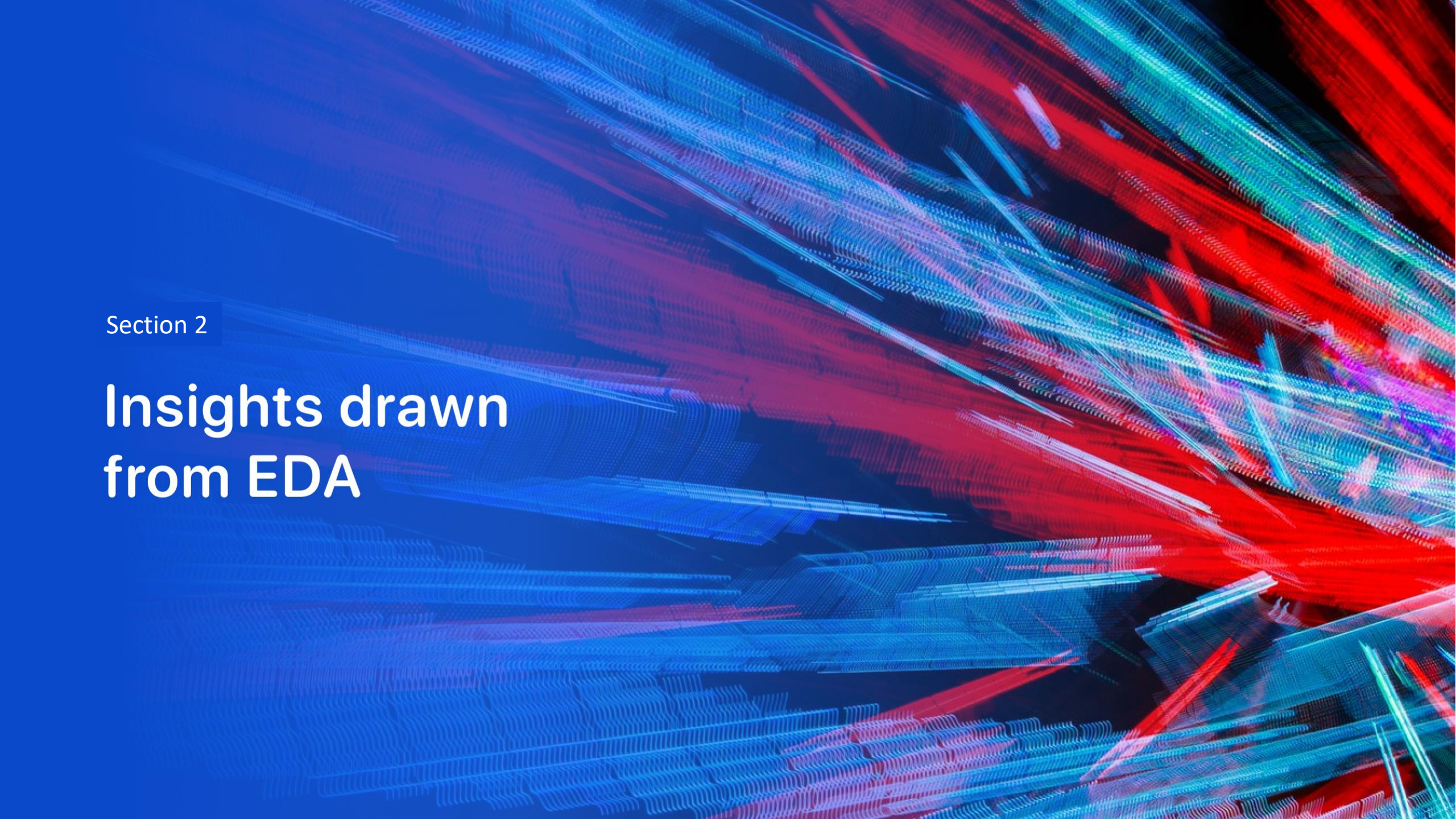
Calculate accuracy of the test data using .score() for all models

Assess the confusion matrix for all models

Identify the best model using Jaccard_Score, F1_Score, and Accuracy

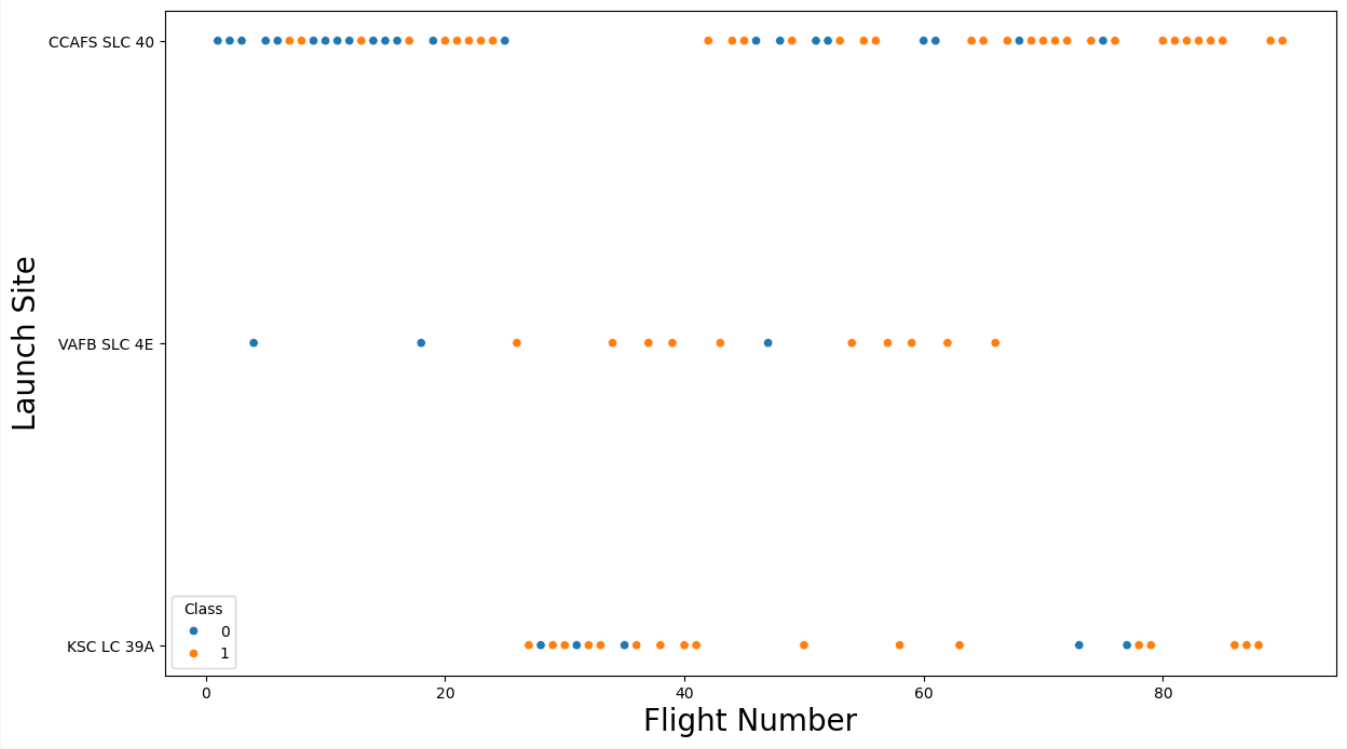
Results

- Exploratory data analysis results somewhat show that as flight number increases, successful landings somewhat do as well
- All the various machine learning methods were able to predict landing successful with an accuracy above 80%


Section 2
Insights drawn from EDA

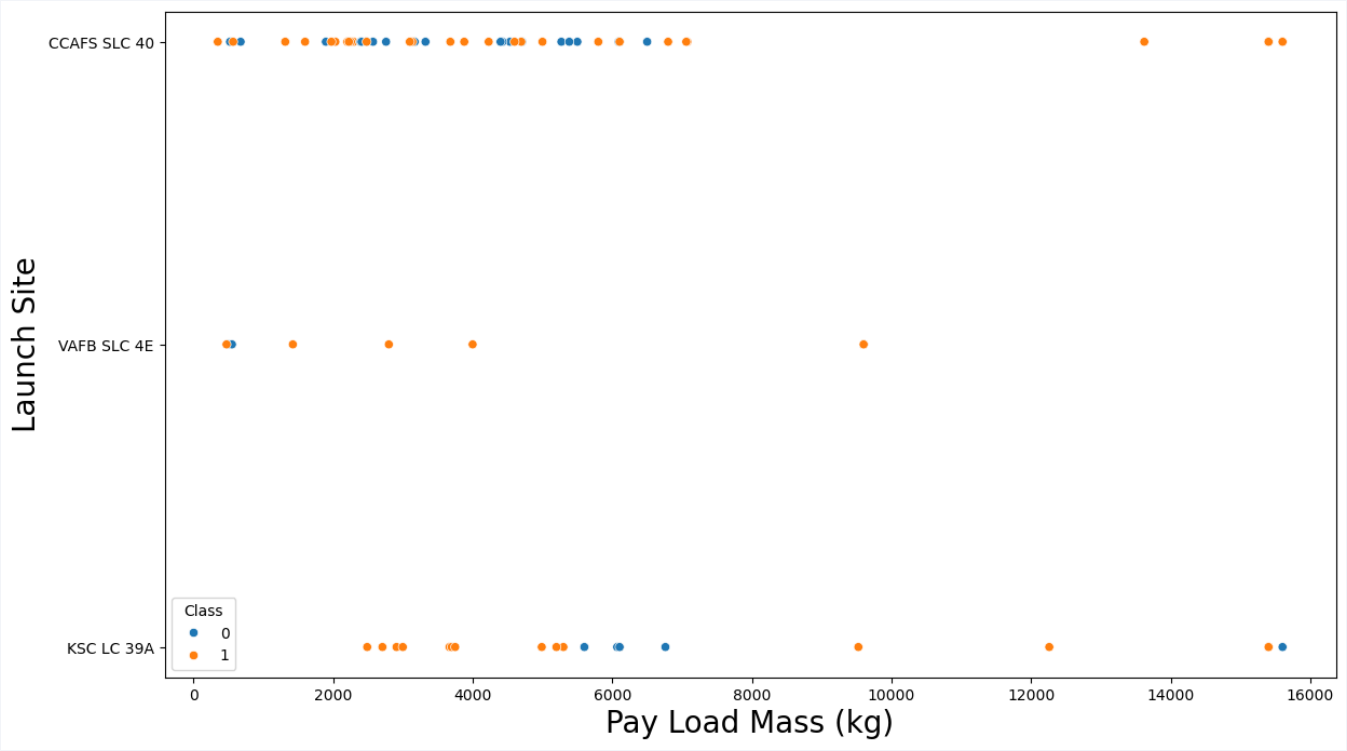
Flight Number vs. Launch Site

- As the number of launches increased, so did the rate of success
- CCAFS SLC 40 had the most successful launches



Payload vs. Launch Site

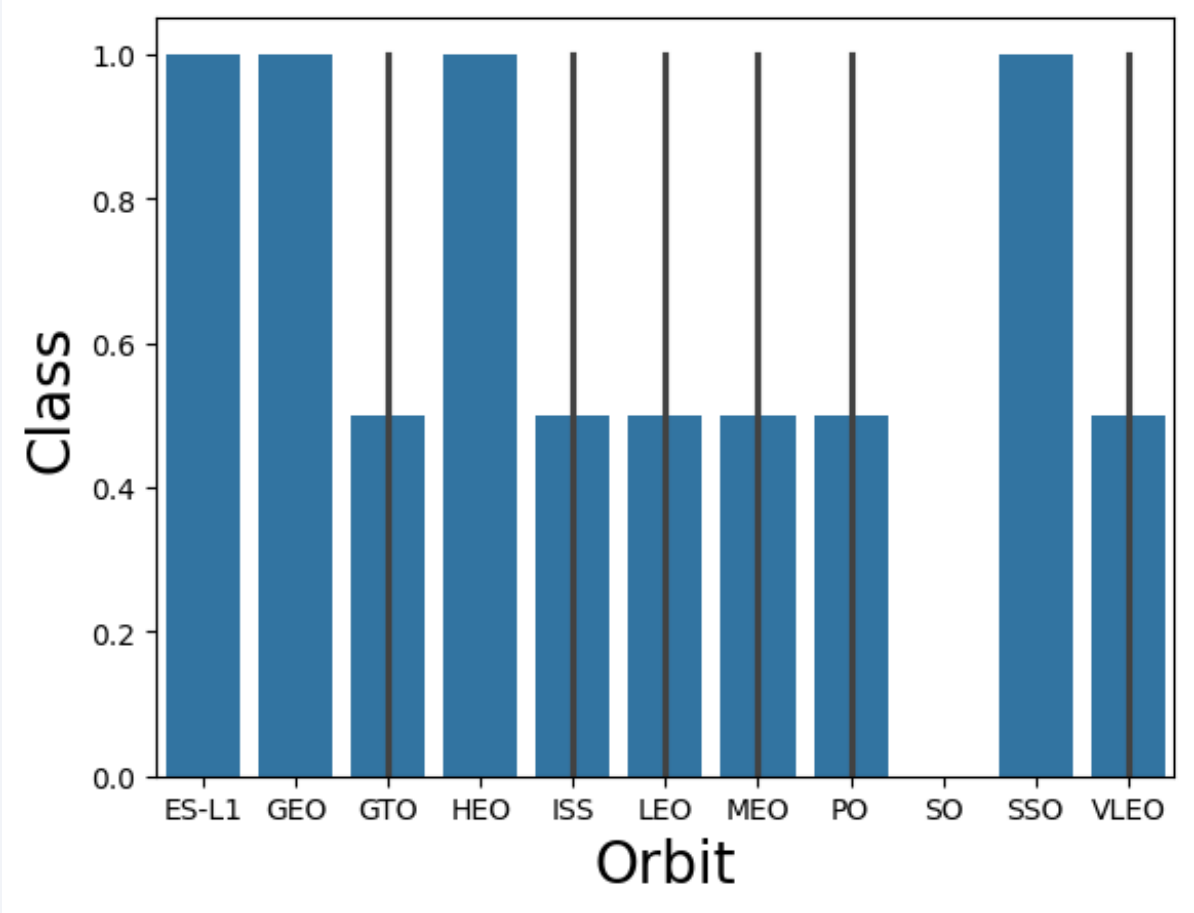
- VAFB SLC 4E is the only launch site to not have rockets that carry a payload more than 10000 kg, and the other two sites only did so thrice



Success Rate vs. Orbit Type

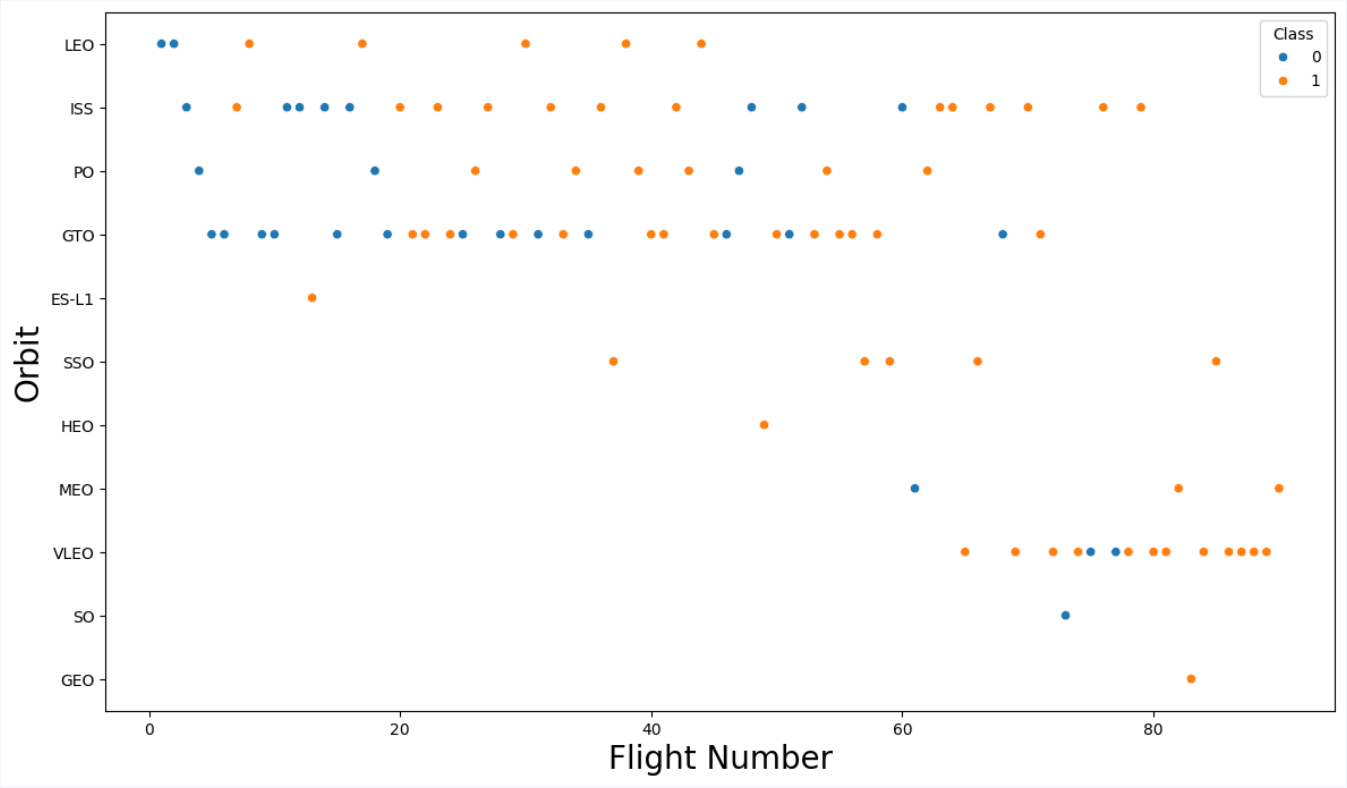
The four orbits with the highest success rates with a class of 1.0 are:

- ES-L1
- GEO
- HEO
- SSO



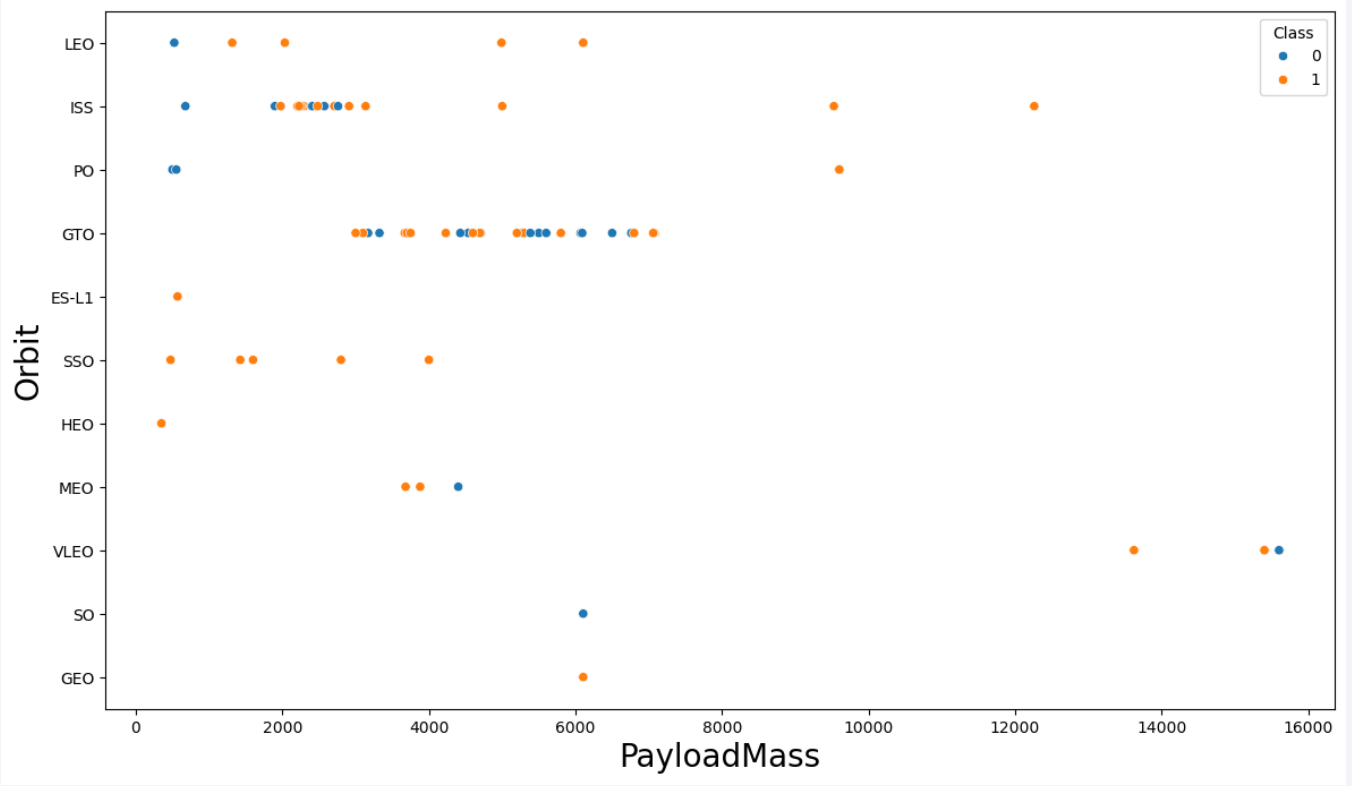
Flight Number vs. Orbit Type

- The orbit LEO's success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when GTO is in orbit.



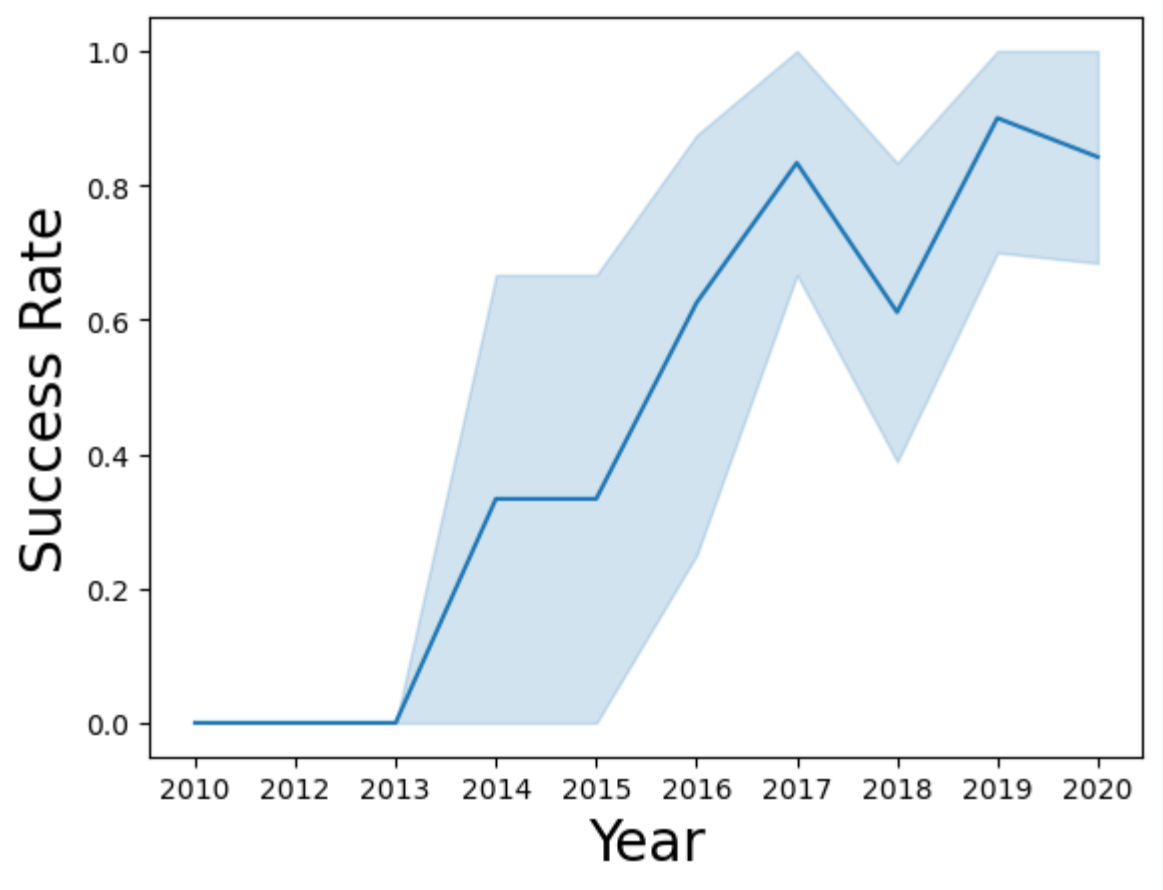
Payload vs. Orbit Type

- Even with heavy payloads, orbits Polar, LEO, and ISS have the most successful landing rates, or positive landing rates
- However, for GTO we cannot distinguish this well as both positive landing rate and negative landing (unsuccessful mission) are seen



Launch Success Yearly Trend

- 2019 saw the highest success rate
- Overall, the rate of success has improved drastically since 2013



All Launch Site Names

Florida

- Cape Canaveral Space Launch Complex 40 (CCAFS SLC 40)
- Kennedy Space Center Launch Complex 39A (KSC LC 39A)

California

- Vandenberg Space Launch Complex 4 (VAFB SLC 4E)

- Launch site CCAFS LC 40 was later renamed CCAFS SLC 40

Launch Site Names Begin with 'CCA'

- Using SQL, 5 records come up

Display 5 records where launch sites begin with the string 'CCA'

```
%sql SELECT * FROM SPACEXTBL WHERE launch_site LIKE 'CCA%' LIMIT 5;
```

```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- NASA's boosters carried a whooping total of 45596kg

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%%sql
SELECT SUM(payload_mass__kg_) AS total_payload_mass
FROM SPACEXTBL
WHERE customer = 'NASA (CRS)';
```

```
 * sqlite:///my_data1.db
Done.
```

total_payload_mass
45596

Average Payload Mass by F9 v1.1

- Whereas F9 v1.1 carried an average of 2928.4kg

Display average payload mass carried by booster version F9 v1.1

```
%%sql
SELECT AVG(payload_mass__kg_) AS average_payload_mass
FROM SPACEXTBL
WHERE booster_version = 'F9 v1.1';
```

```
 * sqlite:///my_data1.db
Done.
```

average_payload_mass
2928.4

First Successful Ground Landing Date

- 2015-12-22 was the day of the 1st successful ground landing

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
%%sql
SELECT MIN(date) AS first_successful_landing_date
FROM SPACEXTBL
WHERE landing_outcome = 'Success (ground pad)';
```

```
 * sqlite:///my_data1.db
Done.
```

first_successful_landing_date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Not all boosters are successful, but these 4 did it with a payload between 4000 and 6000kg:

- F9 FT B1022
- F9 FT B1026
- F9 FT B1021.2
- F9 FT B1031.2

```
%%sql
SELECT booster_version
FROM SPACEXTBL
WHERE landing_outcome = 'Success (drone ship)'
  AND payload_mass__kg_ > 4000
  AND payload_mass__kg_ < 6000;
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- There were a total of 100 successful missions and 1 failed mission

List the total number of successful and failure mission outcomes

```
%%sql
SELECT mission_outcome, COUNT(*) AS total_count
FROM SPACEXTBL
GROUP BY mission_outcome;
```

```
 * sqlite:///my_data1.db
Done.
```

Mission_Outcome	total_count
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters that Carried the Maximum Payload

These boosters went above, but not quite beyond:

- F9 B5 B1048.4
- F9 B5 B1049.4
- F9 B5 B1051.3
- F9 B5 B1056.4
- F9 B5 B1048.5
- F9 B5 B1051.4
- F9 B5 B1049.5
- F9 B5 B1060.2
- F9 B5 B1058.3
- F9 B5 B1051.6
- F9 B5 B1060.3
- F9 B5 B1049.7

```
%%sql
SELECT booster_version
FROM SPACEXTBL
WHERE payload_mass__kg_ = (
    SELECT MAX(payload_mass__kg_)
    FROM SPACEXTBL
);
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

* The Falcon 9 can carry a max payload of 22800 kg

2015 Launch Records

- Two failed launches recorded in 2015

```
%%sql
SELECT strftime('%m', Date) AS month,
       Landing_Outcome,
       Booster_Version,
       Launch_Site
FROM SPACEXTBL
WHERE substr(Date, 0, 5) = '2015'
  AND Landing_Outcome LIKE 'Failure%'
  AND Landing_Outcome LIKE '%Drone Ship%';
```

```
 * sqlite:///my_data1.db
Done.
```

month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- There were 8 successful landings and 5 failed landings during the given time period

```
%%sql

SELECT Landing_Outcome, COUNT(*) AS "Count"
FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' and '2017-03-20'
GROUP BY Landing_Outcome
ORDER BY Count DESC
;
```

```
 * sqlite:///my_data1.db
Done.
```

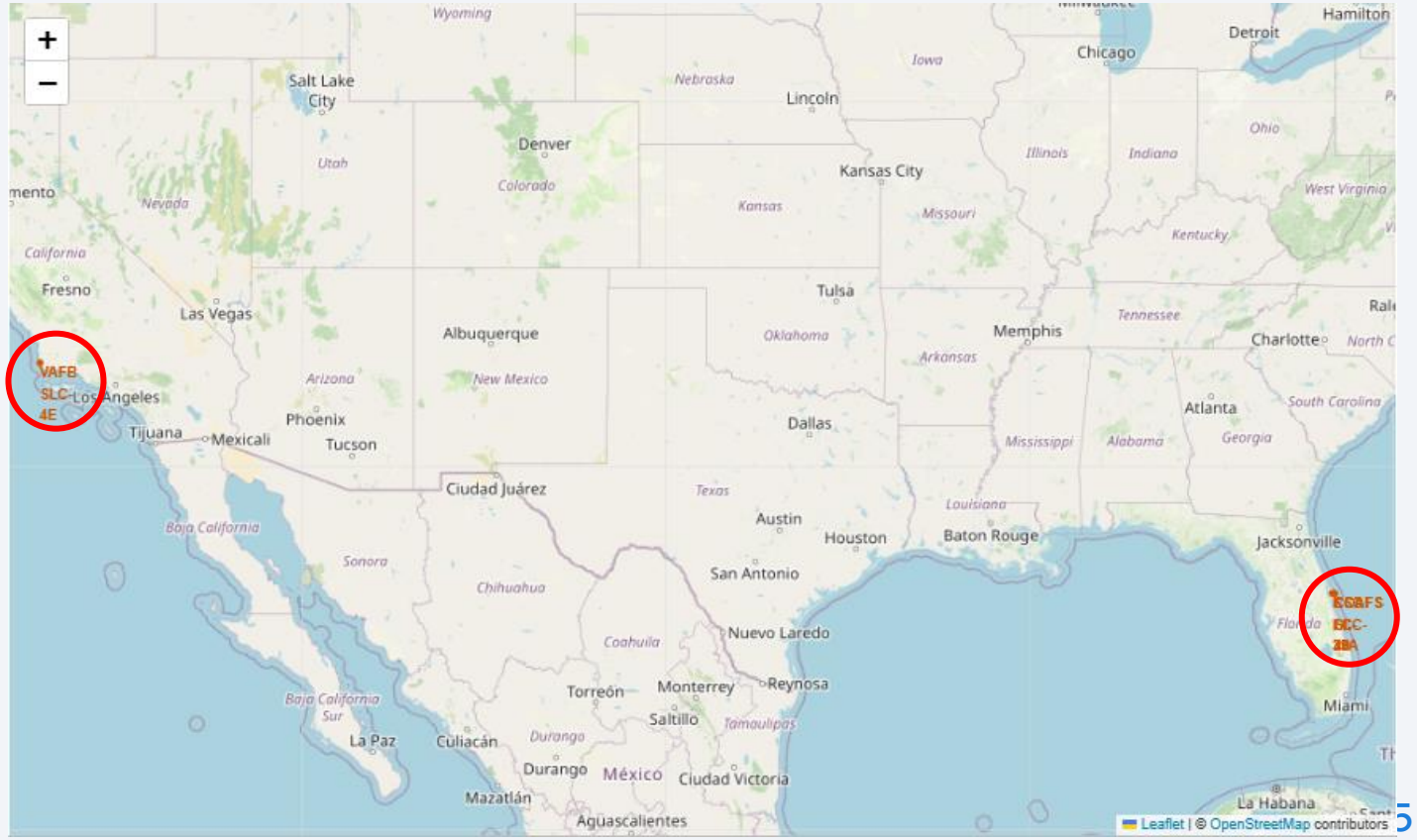
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

Launch Site Locations

- The US launch sites are located on the east and west coasts, allowing for close ocean landings when necessary



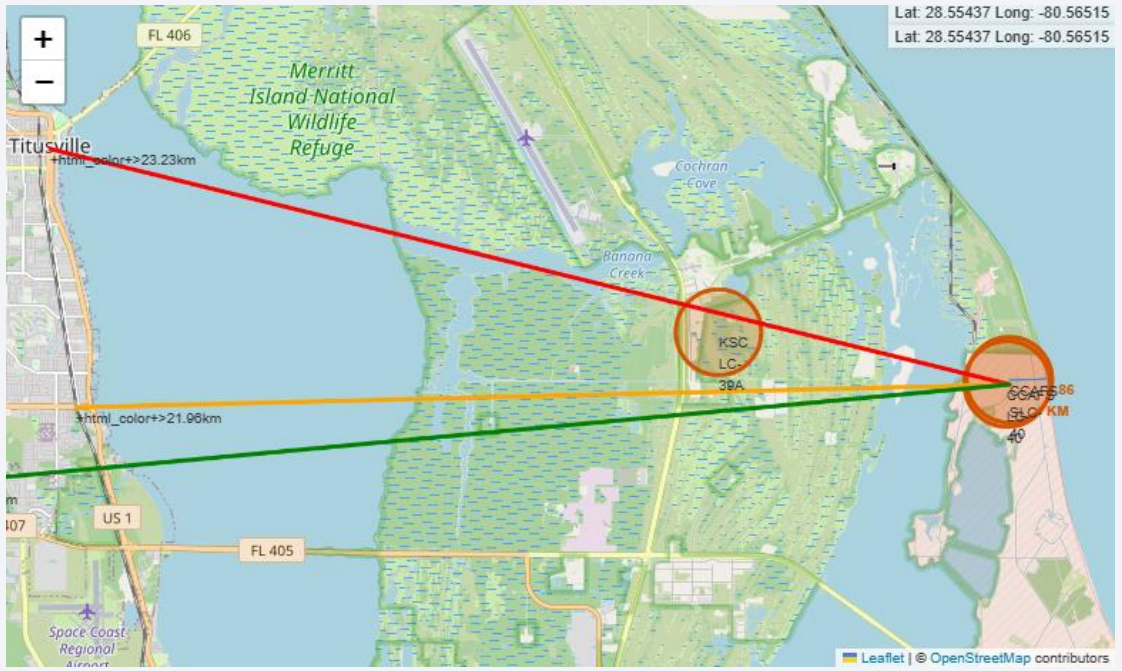
Launch Outcomes

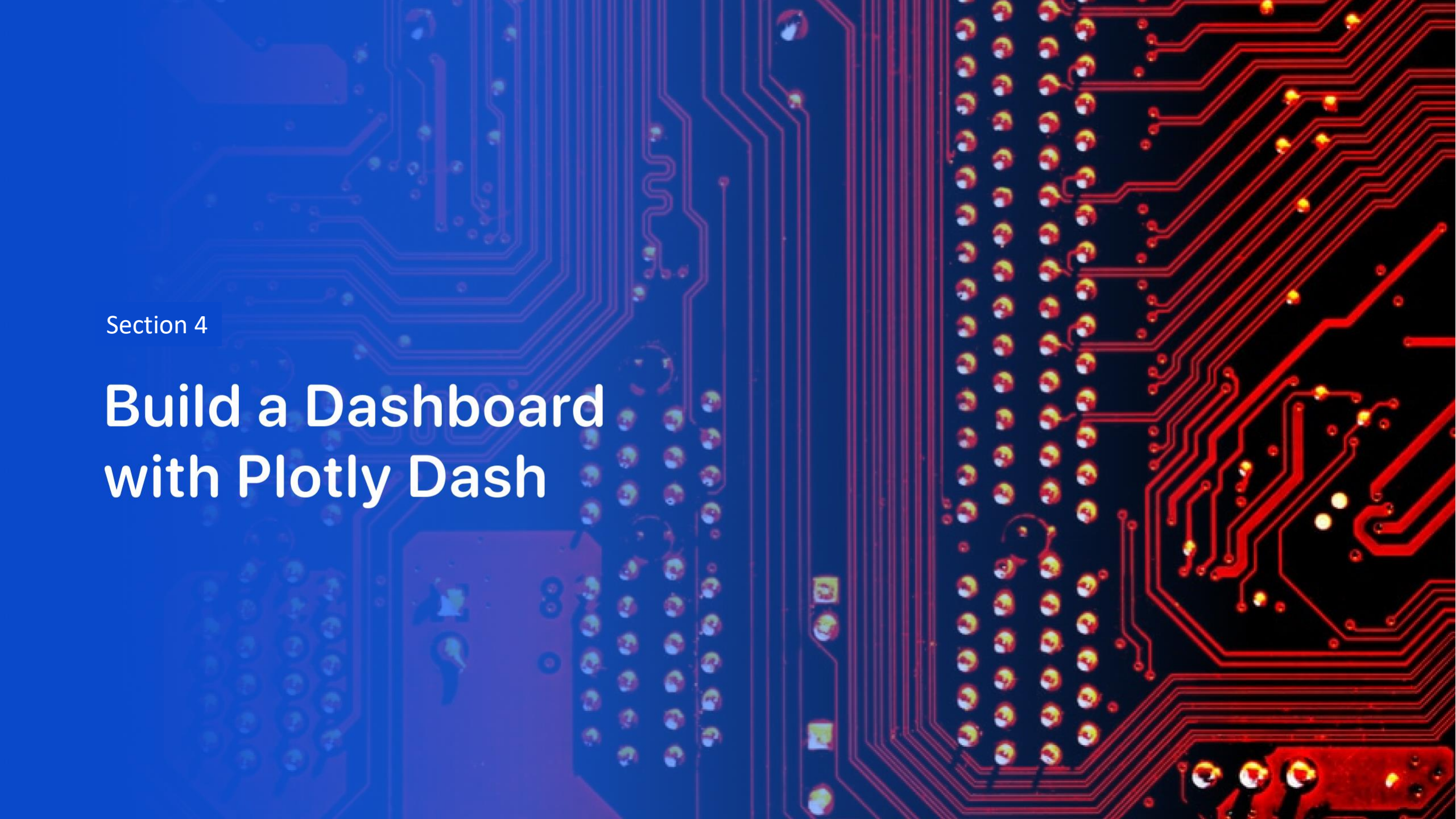
- Successful launches are represented by green markers, while red markers represent failed rocket launches.
- CCAFS SLC 40 has a 3/7 success rate (42.9%)

Surrounding Landmarks

CCAFS SLC 40 is:

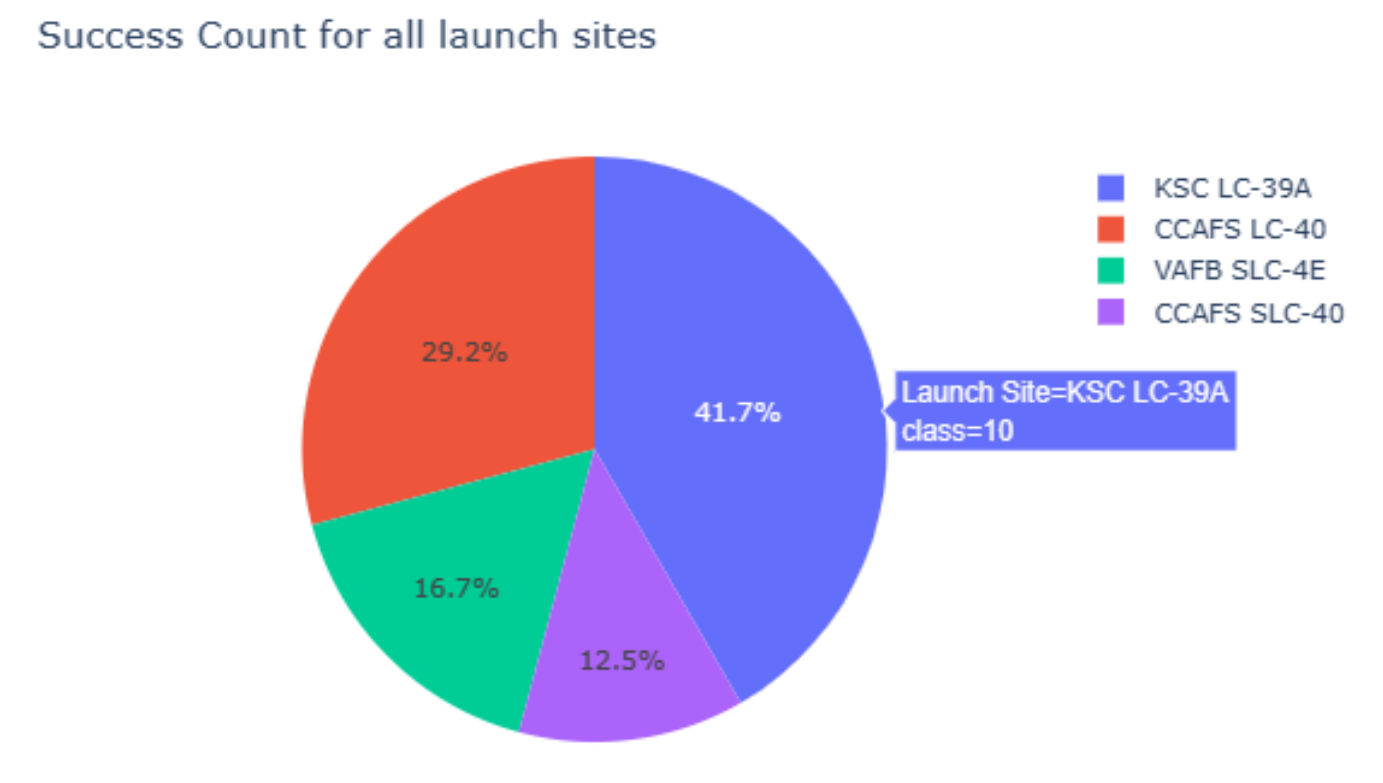
- 21.96 km from the nearest railway
- 26.88 km from the nearest highway
- 0.86 km from the nearest coastline
- 23.23 km from the nearest city




Section 4
Build a Dashboard with Plotly Dash

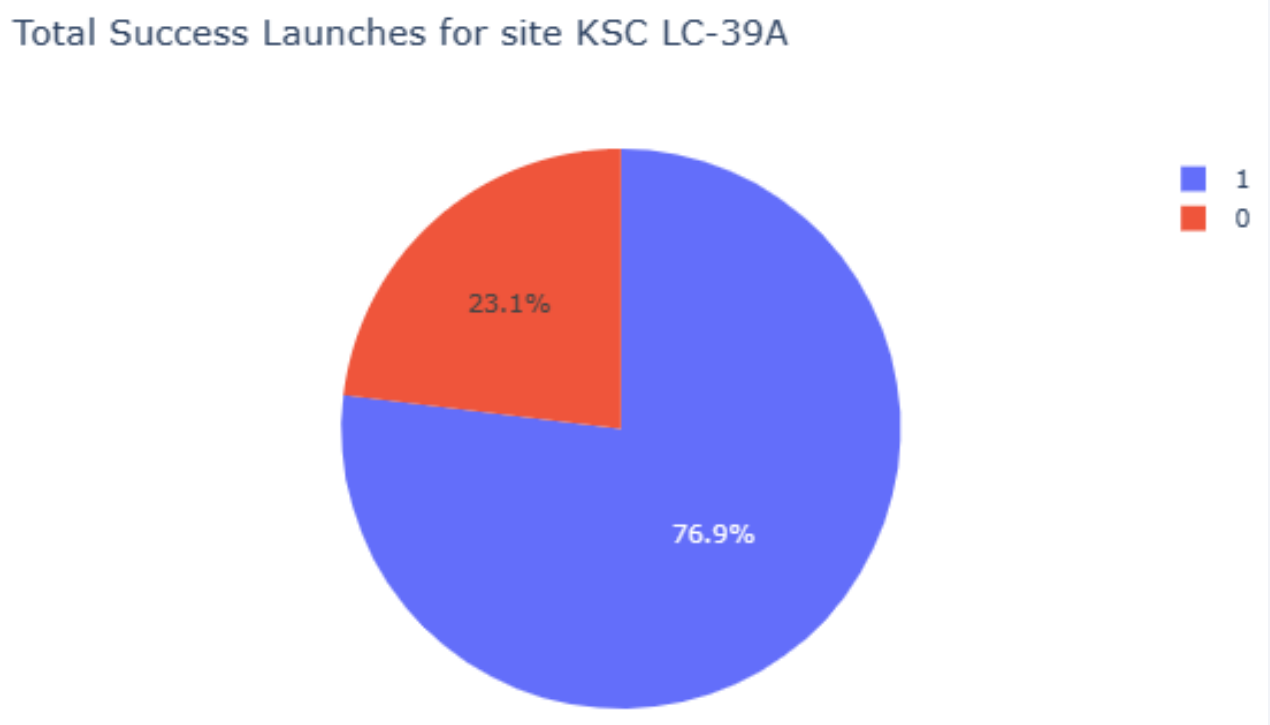
Successful Launches

- With 41.7%, KSC LC 39A has nearly half of the successful launches



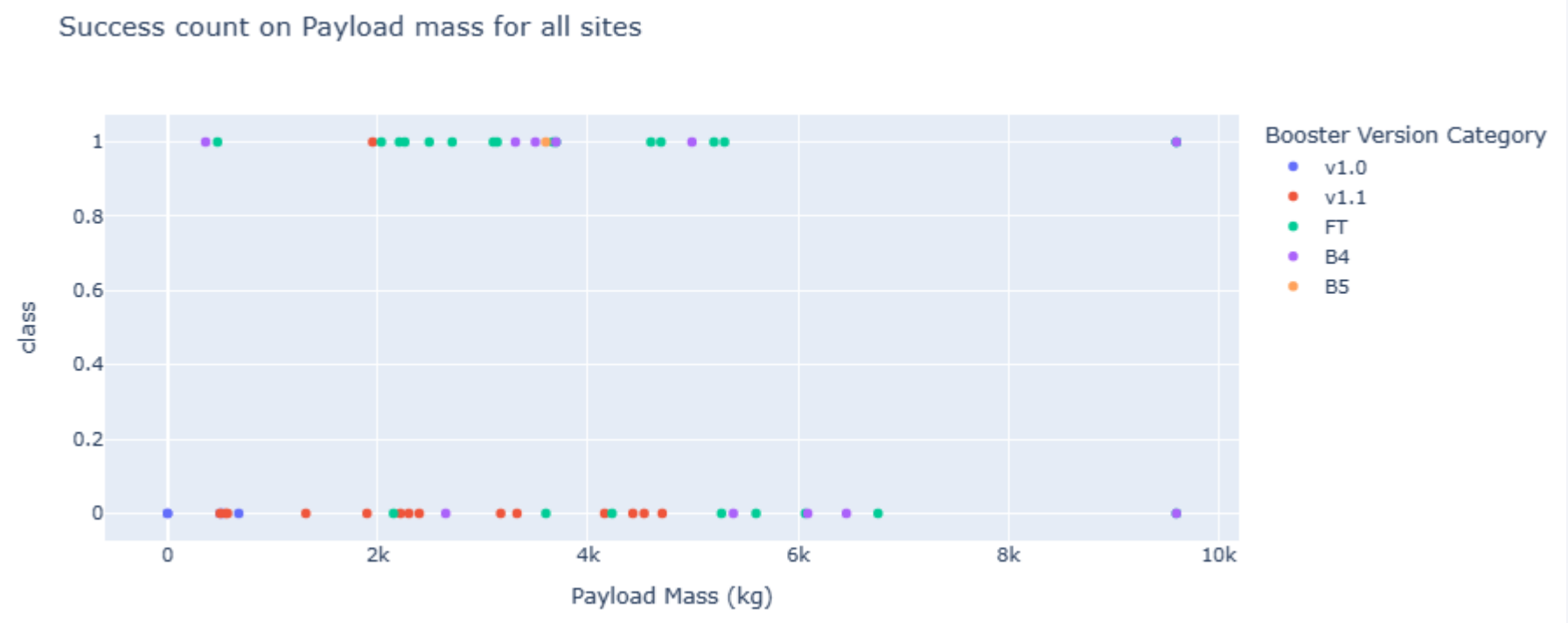
KSC LC 39A's Launch Ratio

- Delving in deeper, KSC LC 39A has a 76.9% success rate



Payload Mass vs. Launch Success for All Sites

- Payloads between 2k and 6k saw the most successes



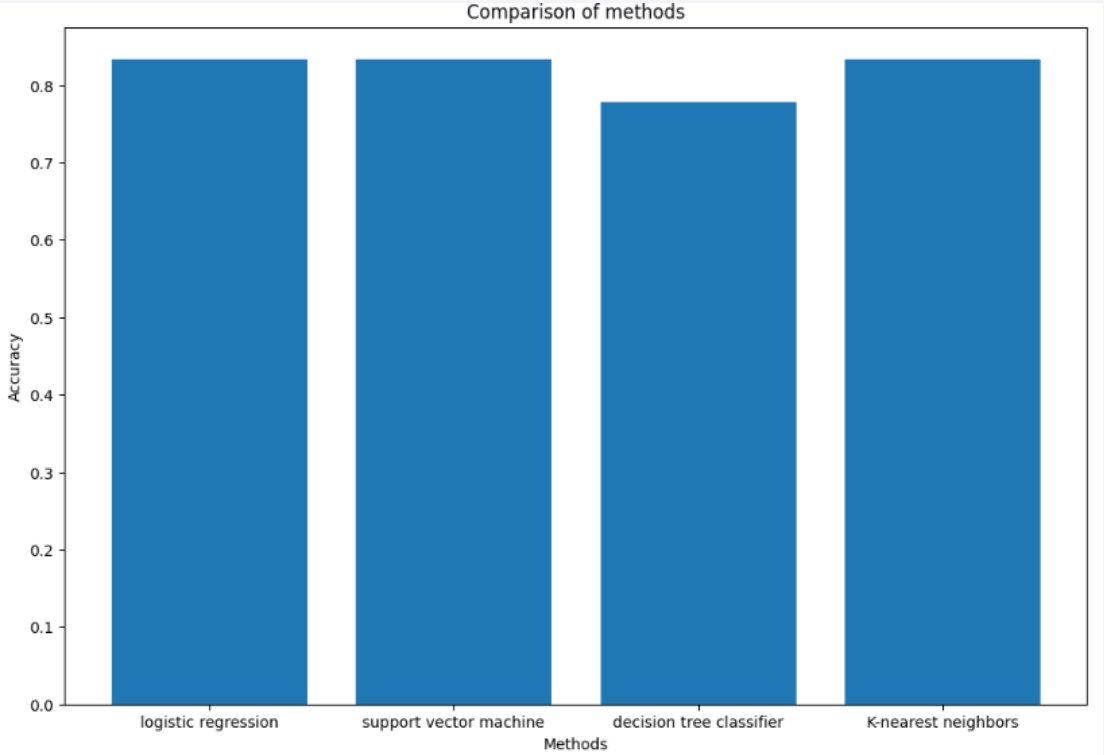
Section 5

Predictive Analysis (Classification)

Classification Accuracy

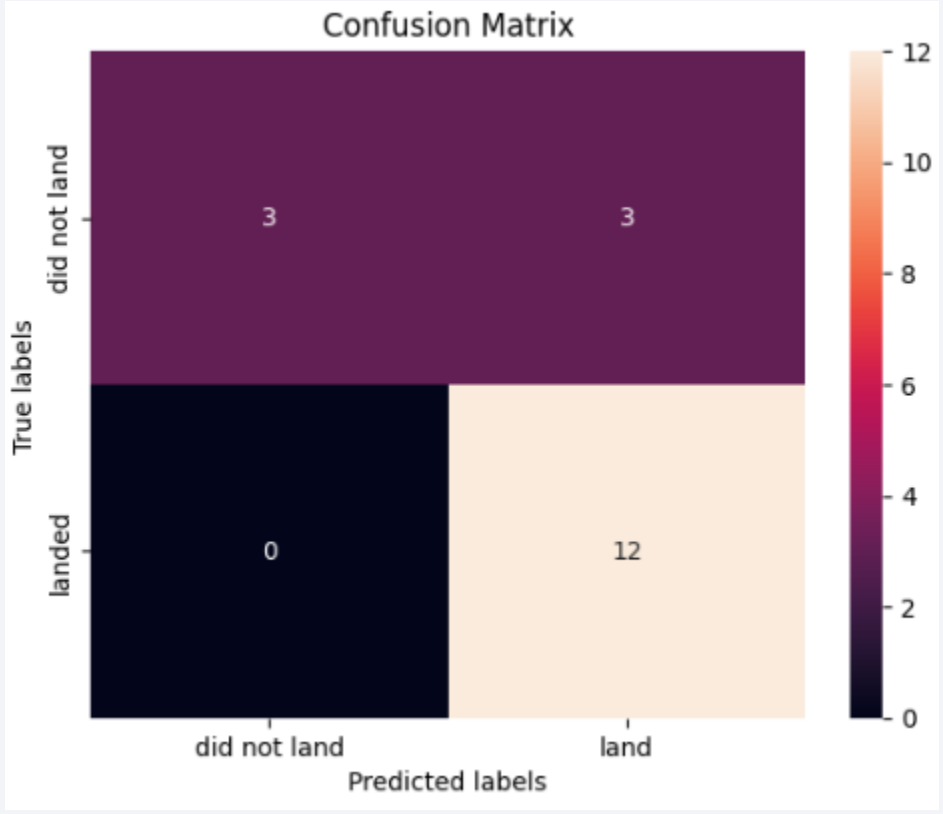
- At 86.07%, the decision tree model has the highest classification accuracy

	Best scores
Logistic regresssion	0.846429
SVM	0.848214
Decision tree	0.860714
KNN	0.848214



Confusion Matrix

- Of the models with confusion matrices, KNN and SVM tied
- The model failed to predict 3 labels accurately



Conclusions

- The goal of this project is to predict the landing outcome of the first stage in a given Falcon 9 launch, to help calculate the launch expenses.
- The mission outcome can be influenced by various Falcon 9 launch features, like payload mass or orbit type.
- The patterns in previous Falcon 9 launch data are learned using various machine learning algorithms in order to generate predictive models for forecasting the outcome of future launches.
- Out of the 4 machine learning algorithms used, the decision tree algorithm yielded the best results with its predictive model.

Appendix

- The goal of this project is to predict the landing outcome of the first stage in a given Falcon 9 launch, to help calculate the launch expenses.

Thank you!